Uncertainties in estimates of cropland area in China: a comparison between an AVHRR-derived dataset and a Landsat TM-derived dataset

Xiangming Xiao\textsuperscript{a,b,*}, Jiyuan Liu\textsuperscript{a}, Dafang Zhuang\textsuperscript{a}, Stephen Frolking\textsuperscript{b}, Stephen Boles\textsuperscript{b}, Bo Xu\textsuperscript{c}, Mingliang Liu\textsuperscript{a}, William Salas\textsuperscript{b}, Berrien Moore III\textsuperscript{b}, Changsheng Li\textsuperscript{b}

\textsuperscript{a}Institute of Geographical Sciences and Natural Resources, Chinese Academy of Sciences, Beijing, China
\textsuperscript{b}Complex Systems Research Center, Institute for the Study of Earth, Oceans and Space, University of New Hampshire, Durham, NH 03824, USA
\textsuperscript{c}Department of Geography, University of New Hampshire, Durham, NH 03824, USA

Accepted 25 July 2002

Abstract

The large uncertainties in estimates of cropland area in China may have significant implications for major cross-cutting themes of global environmental change—food production and trade, water resources, and the carbon and nitrogen cycles. Many earlier studies have indicated significant under-reporting of cropland area in China from official agricultural census statistics datasets. Space-borne remote sensing analyses provide an alternative and independent approach for estimating cropland area in China. In this study, we report estimates of cropland area from the National Land Cover Dataset (NLCD-96) at the 1:100,000 scale, which was generated by a multi-year National Land Cover Project in China through visual interpretation and digitization of Landsat TM images acquired mostly in 1995 and 1996. We compared the NLCD-96 dataset to another land cover dataset at 1-km spatial resolution (the IGBP DIScover dataset version 2.0), which was generated from monthly Advanced Very High Resolution Radiometer (AVHRR)-derived Normalized Difference Vegetation Index (NDVI) from April, 1992 to March, 1993. The data comparison highlighted the limitation and uncertainty of cropland area estimates from the DIScover dataset.

© 2003 Elsevier Science B.V. All rights reserved.

Keywords: Cropland area; China; Landsat TM; AVHRR; Uncertainty

1. Introduction

Traditionally, there have been two approaches to obtain estimates of cropland area in a country: agricultural census statistics and land surveys. Estimates of cropland area in China vary substantially (Fig. 1). The official agricultural census statistics of cropland area in
China are known to be substantial underestimates (Smil, 1995, 1999; Crook, 1993), but have been used in many studies of food security and biogeochemical cycles (Zhang, 1996; Frolking et al., 1999; Brown, 1995; Li et al., 2001). In the book "Who will Feed China?" (Brown, 1995), recent trends in population growth, development, land-use, and food consumption in mainland China were compared with those of Japan, South Korea, and Taiwan. Brown (1995) projected that in another decade or two China’s food demand might far outweigh its ability to produce food, with profound implications for both Chinese and global agricultural trade and food supplies. In response to Brown (1995), there have been intense domestic and international debates about the current status and potential of agriculture in China (Smil, 1995; Chen, 1996; Alexandratos, 1996; Zhang, 1996; Mitchell et al., 1997; Ash and Edmonds, 1998; Heilig, 1999). Mitchell et al. (1997) used an econometric 'World Grain Model' and assumptions about continuing improvements in grain yield to argue that grain production can increase sufficiently in China and around the world for at least the next 20 years. Other studies suggested (1) that China still has great potential for expanding both crop area and crop yields (Zhang, 1996) and (2) that China can feed itself for the next 25 years if it adopts appropriate economic and policy measures, and if it develops its water resources infrastructure (Heilig, 1999). Smil (1995) cautioned that all analyses were compromised by poor quality or uncertain data on major factors, such as cropland area, crop yields, and grain reserves. Uncertainty in estimates of cropland area in China will propagate through any analysis of food security, biogeochemical cycles, and land use and land cover changes.

Space-borne remote sensing technology provides an alternative and independent approach to estimate areal extent and spatial distribution of cropland in a country. Numerous studies have used image data from the Advanced Very High Resolution Radiometer (AVHRR) sensors (at spatial resolutions of 1, 4 and 8 km) to generate maps of land cover at the country- to continental scales over the last few decades. As part of the National Aeronautics and Space Administration (NASA) Earth Observing System Pathfinder Program and the International Geosphere–Biosphere Programme (IGBP) Data and Information System, monthly AVHRR-derived Normalized Difference Vegetation Index (NDVI) data from April, 1992 to March, 1993 were used to generate the global land cover...
The land cover classification scheme of the IGBP DIScover dataset has 17 cover types, including “Cropland” and “Cropland/Other Vegetation Mosaic”. The first version of the global land cover database was completed and released to the public in November, 1997 (IGBP DIScover version 1.2). As an independent estimate of cropland area, the DIScover data set (version 1.2) was compared with FAO agricultural statistics data at the global scale (Ramankutty and Foley, 1998), and agricultural census statistics data in China (Frolking et al., 1999) and in the USA (Hurtt et al., 2001). The comparison between the DIScover and agricultural census data (1990) for China at national and county scales highlighted a large discrepancy in estimates of cropland area (Frolking et al., 1999). After incorporating feedback from the users of the DIScover (version 1.2) database (Brown et al., 1999) and validation exercises (Scepan, 1999; Muchoney et al., 1999), the DIScover version 2.0 was recently released to the public and contains updated land cover and water classes.

In addition to AVHRR data, numerous studies have also used Landsat Thematic Mapper (TM) data at 30-m spatial resolution to map and monitor cropland at landscape to regional scales. After their first large-scale analysis of TM images for China in the early 1990s (Liu, 1996), scientists from the Chinese Academy of Sciences have recently developed a National Land Cover Dataset (NLCD-96) through visual interpretation and digitization of TM images. Most of the source TM images were acquired in 1995/1996. In this study, we report cropland estimates from NLCD-96 at the 1:100,000 scale (Liu et al., 1999), and compare the NLCD-96 dataset to the IGBP DIScover dataset at 1-km spatial resolution (Loveland et al., 2000). This data comparison study at the pixel and national scales will help quantify the uncertainty of cropland area estimates for China from the IGBP DIScover dataset.

2. Data and methods

In this study, we used both the AVHRR-derived IGBP DIScover (version 2.0) dataset and the Landsat TM-derived NLCD-96 dataset for China. As the DIScover dataset was described in Loveland et al. (2000), here we only give a short description of the NLCD-96 dataset. In the late 1990s the Chinese Academy of Sciences organized eight research institutions and about 100 scientists to conduct its second national-scale land cover and land use classification project, using 520 TM images primarily from 1995/1996. The images were geo-referenced and ortho-rectified, using field-collected ground control points and high-resolution digital elevation models, and have an average geoposition error of ± 50 m. Visual interpretation and digitization of TM images at a scale of 1:100,000 was conducted to generate thematic maps of land use and land cover. A hierarchical classification system of 25 land cover classes was used, including two cropland classes (paddy fields and non-flooded cropland). Interpretation of TM images and validation of land cover classification were based on extensive field surveys. For example, more than 7900 field photos were taken using cameras equipped with global positioning system receivers, totaling about 75,000 km of transects across China. The validation results showed that the overall accuracy of the land cover classification approached 98.7%. Additional corrections were made to the NLCD-96 cropland area estimates to account for the fraction of non-cultivated land (e.g., narrow roads and footpaths, small rice paddy levees, irrigation channels) within a polygon of cropland. To account for this fraction of non-cultivated land, the NLCD-96 project developed a stratified, multi-layer sampling design that divided China into 870 sampling zones, and acquired aerial photos for each sampling zone. In the analysis presented below, we use the original NLCD-96 product (uncorrected) to develop a comparison between AVHRR-based and Landsat-TM based land cover products.

The vector Map of Land Cover in China (1:100,000 scale) was converted into a 1-km gridded database that still captures all of the high-resolution land cover information by calculating percent fractional cover within 1-km grid cell (Tang, 2000; Zhang et al., 2000; Liu et al., 2001). We compared the TM-derived 1-km gridded database (NLCD-96) with the AVHRR-derived 1-km gridded IGBP DIScover (version 2.0) dataset. For each 1-km grid cell, the NLCD-96 dataset contains percent fractions of paddy fields and non-flooded croplands, which were added to calculate total cropland area for every 1-km grid cell. The DIScover dataset was provided by the USGS EROS Data Center (EDC). We downloaded the data-
Fig. 2. A comparison of the spatial distribution of cropland at 1-km resolution. (a) Percent paddy fields within a 1-km gridcell from the NLCD-96 dataset; (b) percent non-flooded cropland within a 1-km gridcell from the NLCD-96 dataset; (c) percent cropland (sum of paddy fields and non-flooded cropland) within a 1-km gridcell from the NLCD-96 data; (d) “Cropland” (green color) and “Cropland/Other Vegetation Mosaic” (orange color) from the IGBP DIScover dataset (version 2.0).
set for the Eurasia continent (in Lambert Azimuthal Equal Area Projection optimized for Asia) from the EDC website. The digital version of the China Administration Map of 1995/1996 (from the Chinese National Survey Agency) was used to subset the Eurasia dataset for China.

3. Results

In the hierarchical land cover classification scheme of the NLCD-96 dataset, cropland is sub-classified as either paddy fields (mostly for growing rice, though frequently double-cropped; Fig. 2a) or non-flooded cropland (e.g., dryland farming for wheat, rapeseed, corn and soybean crops; Fig. 2b). It is estimated that there are 0.50 million km² of paddy fields distributed over 1.36 million 1-km grid cells (Fig. 2a), and 1.24 million km² of non-flooded cropland distributed over 2.9 million 1-km grid cells in China (Fig. 2b). By adding both paddy fields and non-flooded cropland together, there is 1.74 million km² total cropland area in China, distributed over 3.6 million 1-km grid cells (Fig. 2c). Approximately 53% of those 3.6 million 1-km grid cells have ≤50% cropland fraction (Fig. 3).

In the DIScover dataset there are approximately 1.8 million pixels of “Cropland” and 0.95 million pixels of “Cropland/Other Vegetation Mosaic” in China (Fig. 2d). Each pixel in the DIScover dataset has an area of 1 km². The DIScover dataset is the result of a per-pixel classification approach (by dominant cover type) and contains little quantitative information of sub-pixel components. However, at 1-km spatial resolution, the land surface is mostly a mixture of various land cover types, including cropland, forest, grassland, soil and water. Some DIScover land cover class descriptions contain a bounding fractional area; for example, “Cropland/Other Vegetation Mosaic” is defined such that ‘no one component comprises more than 60% of the landscape’ (Belward, 1996), implying that “Cropland” pixels should have >60% cropland fraction. In order to obtain an estimate of cropland area from the DIScover dataset, it is necessary to assign a sub-pixel cropland fraction to each 1-km pixel of “Cropland” and “Cropland/Other Vegetation Mosaic”. Spatial variability in sub-pixel cropland fraction within a 1-
A 1-km pixel will introduce uncertainty into a large-scale estimate and assessment of cropland area.

In the simplest fashion, Frolking et al. (1999) assumed a 100% cropland fraction within a “Cropland” pixel and a 50% cropland fraction with a “Cropland/Other Vegetation Mosaic” pixel to obtain an upper-bound estimate. Applying these percent cropland fractions to DIScover generated an estimate of total cropland area in China of 2.26 million km². There are few studies that have quantified percent cropland fraction within a 1-km pixel at the landscape scale, using both fine-resolution images (e.g., Landsat TM) and coarse-resolution images (e.g., AVHRR). Xiao et al. (2002) conducted a land cover classification for the Nanjing area (22,500 km²) in Jiangsu Province, China, using a 1996 Landsat TM image at 30-m spatial resolution and multi-temporal (3/1999–5/1999) 10-day composite data from VEGETATION (VGT) sensor. The VGT sensor was designed for global vegetation observation with four spectral bands at 1-km resolution, and was launched in March, 1998 on the SPOT 4 satellite. Xiao et al. (2002) used an unsupervised classification on the Landsat TM image and the VGT images to classify the landscape into 17 land-cover classes (the IGBP land cover classification scheme in the DIScover dataset), including “Cropland” and “Cropland/Other Vegetation Mosaic” types. Based on the co-registered VGT and Landsat TM classified images for the Nanjing area, Xiao et al. (2002) estimated that each 1-km VGT pixel classified as “Cropland” had an average fractional cropland cover of approximately 60%, and each 1-km VGT pixel classified as “Cropland/Other Vegetation Mosaic” had an average fractional cropland cover of approximately 40% (Xiao et al., 2002). We applied these 60% and 40% fractional cropland estimates within a 1-km pixel to the DIScover dataset to obtain an estimate of total cropland area in China of 1.45 million km².

The DIScover map of cropland distribution looks very similar to the NLCD-96 map of cropland distribution, with concentrations of cropland in the broad river valleys of northeastern China, the North China Plain, and the Sichuan basin, and little cropland along the northern border or in the western half of the country (Fig. 2). In hilly southern China, however,
many pixels in DIScover were classified as “Cropland” while NLCD-96 fractional cropland areas were generally < 40% (Fig. 2). The availability of NLCD-96 at 1-km resolution (Fig. 2) enables us to calculate percent fractional cropland area within a 1-km pixel for all the “Cropland” and “Cropland/Other Vegetation Mosaic” pixels in DIScover, respectively. After co-registering the two datasets, we performed a spatial query for all “Cropland” and “Cropland/Other Vegetation Mosaic” pixels of the DIScover dataset in relation to fractional cropland cover within a 1-km pixel from the NLCD-96 dataset. The resulting histograms were rather flat (Fig. 4), indicating that land classified as “Cropland” or “Cropland/Other Vegetation Mosaic” in the DIScover dataset can have any fractional cropland area within a 1-km pixel (as determined by NLCD-96) with almost equal probability. The cumulative frequency curves show that about half the DIScover “Cropland” pixels have ≤ 60% fractional cropland cover, and about half of the DIScover “Cropland/Other Vegetation Mosaic” pixels have ≤ 20% fractional cropland cover within a 1-km pixel (Fig. 4). Fourteen percent of the “Cropland” pixels and 26% of the “Cropland/Other Vegetation Mosaic” pixels had no cropland area in the NLCD-96 dataset (Fig. 4). The “Cropland” pixels in the DIScover dataset contain 0.98 million km² of the cropland area of the NLCD-96 dataset, while the “Cropland/Other Vegetation Mosaic” pixels contain 0.30 million km² of the cropland area of the NLCD-96 dataset. The sum of 1.28 million km² of cropland accounts for 74% of the 1.74 million km² total cropland area in the NLCD-96 dataset. On average, “Cropland” and “Cropland/Other Vegetation Mosaic” pixels in the DIScover dataset have means of 55% and 32% fractional cropland cover within a 1-km pixel, respectively.

4. Discussion

The comparison between the NLCD-96 and DIScover datasets at the pixel and national scales highlight the limitation of the DIScover dataset, mostly attributed to its coarse spatial resolution (1 km) and the per-pixel classification approach (by dominant land cover types), which may have been particularly problematic in the hilly regions of southern China. Caution should be employed when using the DIScover dataset for large-scale simulations of biogeochemical cycles and land use and land cover change. The NLCD-96 dataset represents significant progress in land cover classification in China, however, it still has significant uncertainties. The original 1:100,000 scale Map of Land Cover in China generated in the NLCD project (and used in this study) did not consider the fraction of non-cultivated land within a polygon of cropland, and therefore, its estimate of 1.74 million km² cropland area for China is too high. The interpretation of aerial photos showed that the fraction of non-cultivated land in a cropland polygon ranges between 20% and 30% for most of the 816 sampling zones where the original 1:100,000 scale Map of Land Cover in China indicates that there is cropland (Fig. 5). Applying each zone’s fraction of non-cultivated land to the 1:100,000 Map of Land Cover in China, NLCD-96 generated an estimate of 1.35 million km² total cropland area in China.

Similar magnitude estimates of cropland area in China were obtained by some other land surveys. For example, the agricultural land resource survey in the early 1980s estimated about 1.40 million km² cropland area in China (Fig. 1; Liu, 1996). Cultivated land area in China was estimated to be approximately 1.36 million km² (including cropland area in Taiwan, Hong Kong and Macao), according to a 1:1,000,000 Land Use Map of China (Wu, 1990; Wu and Guo, 1994) that was generated from extensive field surveys, and interpretation of aerial photos and Landsat images during the late 1970s and the early 1980s. Estimates of cropland area from land surveys and remote sensing approaches are substantially higher than the official agricultural census statistics (Fig. 1; State Statistical Bureau, 1995). The issue of under-reporting of cropland area in the official agricultural census statistics has been documented in a number of studies (e.g., Seto et al., 2000; Smil, 1999), and they have suggested several explanations for the under-reporting, including institutional bias, grain production quotas, and cropland taxes. Yuyun and Zheng (2000) report that crop area estimates for China were accurate in 1960, but since that time only cropland losses have been tallied, while cropland area gains, generally from conversion of more marginal lands, have not. The result has been a fairly steady decline in total cropland area reported since 1960. Two reasons are given for this under-reporting of cropland area: reducing tax burdens and...
artificially inflating crop productivity numbers (Yuyun and Zheng, 2000). Larger estimates of cropland area in China from the remote sensing and land survey approaches than the official agricultural census statistics (Fig. 1) indicate that average yield per unit area in China may be lower than reported, and thus there is additional potential to raise crop production through improved management.

The approach of visual interpretation and digitization of TM images in the NLCD-96 project did not take full advantage of either the 30-m spatial resolution of the imagery or recent advances in digital processing and classification of TM images. To improve estimates of cropland area in China, future large-scale Landsat-based image analysis projects in China should explore digital classification of TM images at 30-m spatial resolution, as was done in a national-scale land cover study for the conterminous United States (Vogelmann et al., 2001). A global archive of ortho-rectified Landsat data is available in early 2002, including Multispectral Scanner (MSS) images from the late 1970s and TM images from the late 1980s and early 1990s (Kalluri et al., 2000), which can be used to quantify land-use and land-cover changes in China. Even at 30-m resolution, however, a pixel classified as cropland can have a fraction of its area that is non-cropland, particularly in many regions of China where a typical agricultural field will occupy only a few tenths of a hectare. For any map-based or image-based dataset, there is uncertainty as to how much of the land area designated as cropland is actually used to grow crops. Some non-planted land within a cropland pixel such as a rice paddy irrigation canal is clearly agricultural infrastructure, and for many analyses, should probably be counted as cropland. Other non-planted land within a cropland pixel, such as a road or footpath, has a more general infrastructure function and should probably not be counted as cropland. For some analyses, such as potential carbon sequestration in cropland soils, only actual cultivated cropland area should be counted, and not other agricultural infrastructure land. For other analyses, such as land area available for development, all necessary infrastructure land should be included in cropland area totals. Aerial photography used in conjunction with the NLCD-96 analysis showed that in China (at 1:100,000 scale) about 20–30% of land designated as cropland is not used directly to grow crops (Fig. 5). At coarser resolution (e.g., DIScover),
this fraction is likely to be larger, while at finer resolution (e.g., native resolution of Landsat imagery) the fraction is likely to be smaller. To quantify non-cropland fractions within 30-m resolution TM image pixels, it will be necessary to acquire and analyze aerial photos and/or very high-resolution space-borne images (e.g., 1–4 m spatial resolution IKONOS images) under a stratified and multi-layer sampling design.

Acknowledgements

This study was part of joint research effort between the Institute of Geographical Sciences and Natural Resources, Chinese Academy of Sciences, and the Institute for the Study of Earth, Oceans and Space, University of New Hampshire. It was supported in part by the National Land Cover Characterization Project from the Chinese Academy of Sciences, a grant (NAG5-6137, NAG5-10135) from the U.S. NASA Earth Observing Systems (EOS) Interdisciplinary Study Program, and a grant (NAG5-7631) from the U.S. Multi-Agency Joint Program on Terrestrial Ecology and Global Change (TECO).

References


